

An Analysis of Corrupt Level-2 Jason-2/OSTM Data at NOAA/CLASS

John Lillibridge
NOAA Jason-2 Project Scientist
2009-06-08

1. Introduction

Many months after the launch of Jason-2, we became aware that some of the data products going to NOAA/CLASS had been corrupted during their transfer to the archive. The source of the corruption was isolated to a specific section of code in the FTP module (NOAAServer service) running on the Jason-2 FileManager machine at OSDPD/ESPC. The root cause of the corruption is not completely understood, but was associated with a combination of heavy processing or I/O loads on the FileManager, which resulted in a specific type of error message being embedded in the data file during ftp transfer (apparently this could happen either inbound from the SOCC NOAAServer, or outbound to DDS & CLASS). The error messages which corrupted the files always contained the string 'chown' which helped to isolate the portion of the code where a system call was being made to perform a 'chown -R' command. This section of code was reworked to utilize an internal chown application call, with additional error checking on user/group IDs and on file sizes before & after transfer. This code was implemented on the ESPC Jason-2 FileManager machines on 2009-05-12 at 21:30 UTC, and no further cases of data corruption have been observed.

Unfortunately, data corruption due to this bug may have affected ANY of the 41 data types (not counting XML metadata) which are archived at CLASS. A thorough analysis of the Jason-2 data holdings at CLASS has therefore been undertaken to identify any files corrupted by this 'chown' bug. Thus far we have not observed any corruption in auxiliary/ancillary data files, other than in the ECMWF meteorological grids (filenames beginning with SMM...) which are only available to the four project partners.

More importantly, corruption HAS been observed in the most scientifically useful level-2 data: the near real-time OGDR family (OGDR, OGDR-BUFR, OGDR-SSHA), the interim IGDR family (IGDR, S-IGDR, IGDR-SSHA) and the final high-precision GDR family (GDR, S-GDR, GDR-SSHA). We have concentrated our immediate short-term analysis on this data, since these are widely available to end users: initially to the OSTST, and ultimately to the general public. The complete analysis is still ongoing; this report documents our findings regarding the level-2 products.

Files affected by corruption are identified by comparing their size at CLASS with the same file's size at the CNES/AVISO archive. The corrupt files are then downloaded from CLASS to confirm that they contain the 'chown' error message. The uncorrupt version of the files at CNES are then downloaded and resent to CLASS to replace the corrupt version.

Thus far all cycles of IGDR data have been analyzed, from launch until the software bug was repaired: cycles 998, 999, and 0-31. Similarly, all cycles of GDR data for cycles 0-23 have been analyzed, which is all the data received before the bug fix was implemented. Finally the OGDR data from cycles 0-31 have also been analyzed. The following sections present a detailed list of the specific IGDR, GDR, and OGDR files affected by this corruption.

2. IGDR Data Corruption

- a. In the process of analyzing the CLASS IGDR data for corruption, we discovered that the IGDR, S-IGDR, and IGDR-SSHA data from 10 passes (30 files total) did not match the file size of their CNES counterpart, but did NOT contain the ‘chown’ string. We are currently investigating the cause of this mismatch, but it appears to be associated with a reprocessing at CNES a few days after the original files were received by CLASS. The CLASS version of these 30 files is smaller (sometimes much smaller) than the final CNES files, even though the filenames are exactly the same. The global NetCDF ‘history’ attribute shows that the smaller CLASS files were created 2-3 days before the CNES version. The following files at CLASS have been replaced with the CNES version as part of this effort. ‘File Size’ indicates the correct size for the replacement files from CNES, and ‘CLASS Availability’ shows when the files were replaced at CLASS and available to end users:

Filename	Files Size	CLASS Availability
JA2 IPN 2PTP014 246 20081127 123332 20081127 132945	5037808	2009-05-27 20:27:41
JA2 IPN 2PTP014 248 20081127 142558 20081127 152211	5054852	2009-05-27 20:27:42
JA2 IPN 2PTP015 013 20081128 081403 20081128 091016	5099288	2009-05-27 20:27:40
JA2 IPN 2PTP015 015 20081128 100628 20081128 110241	5108604	2009-05-28 14:14:22
JA2 IPN 2PTP015 056 20081130 003116 20081130 012729	5031660	2009-05-28 14:30:09
JA2 IPN 2PTP015 059 20081130 031955 20081130 041608	5065568	2009-05-28 14:30:08
JA2 IPN 2PTP015 068 20081130 114551 20081130 124204	4979496	2009-05-28 14:30:07
JA2 IPN 2PTP015 070 20081130 133817 20081130 143430	4944188	2009-05-28 14:30:11
JA2 IPN 2PTP015 077 20081130 201147 20081130 210800	4959652	2009-05-28 14:30:05
JA2 IPN 2PTP015 080 20081130 230026 20081130 235638	5071716	2009-05-28 14:30:10
JA2 IPR 2PTP014 246 20081127 123332 20081127 132945	258748	2009-05-28 14:30:12
JA2 IPR 2PTP014 248 20081127 142558 20081127 152211	259608	2009-05-28 14:30:15
JA2 IPR 2PTP015 013 20081128 081403 20081128 091016	261748	2009-05-28 14:30:16
JA2 IPR 2PTP015 015 20081128 100628 20081128 110241	262212	2009-05-28 14:30:17
JA2 IPR 2PTP015 056 20081130 003116 20081130 012729	258448	2009-05-28 14:30:18
JA2 IPR 2PTP015 059 20081130 031955 20081130 041608	260112	2009-05-28 14:30:19
JA2 IPR 2PTP015 068 20081130 114551 20081130 124204	255912	2009-05-28 14:30:20
JA2 IPR 2PTP015 070 20081130 133817 20081130 143430	254208	2009-05-28 14:30:21
JA2 IPR 2PTP015 077 20081130 201147 20081130 210800	254972	2009-05-28 14:30:22
JA2 IPR 2PTP015 080 20081130 230026 20081130 235638	260412	2009-05-28 14:30:23
JA2 IPS 2PTP014 246 20081127 123332 20081127 132945	35483860	2009-05-28 14:35:20
JA2 IPS 2PTP014 248 20081127 142558 20081127 152211	35604260	2009-05-28 14:35:21
JA2 IPS 2PTP015 013 20081128 081403 20081128 091016	35921100	2009-05-28 14:35:22
JA2 IPS 2PTP015 015 20081128 100628 20081128 110241	35986780	2009-05-28 14:35:23
JA2 IPS 2PTP015 056 20081130 003116 20081130 012729	35440136	2009-05-28 14:35:24
JA2 IPS 2PTP015 059 20081130 031955 20081130 041608	35680712	2009-05-28 14:35:24
JA2 IPS 2PTP015 068 20081130 114551 20081130 124204	35068576	2009-05-28 14:35:25
JA2 IPS 2PTP015 070 20081130 133817 20081130 143430	34817228	2009-05-28 14:35:26
JA2 IPS 2PTP015 077 20081130 201147 20081130 210800	34926632	2009-05-28 14:35:36
JA2 IPS 2PTP015 080 20081130 230026 20081130 235638	35724436	2009-05-28 14:35:37

- b. In addition to the 30 files above, another 72 files from the IGDR family were found to be corrupted by the ‘chown bug’ and consequently had a file size mismatch between the CLASS and CNES versions. The vast majority of the corrupt files were of the larger S-IGDR type. Only 3 IGDRs and 1 IGDR-SSHA file were affected by file

corruption. The following CLASS files have been replaced with the uncorrupted CNES version. (Note the change from version 'T' to 'c' during cycle-015):

Filename	Files Size	CLASS Availability
JA2 IPN 2PTP000 119 20080706 175453 20080706 185106	4601300	2009-05-28 14:30:03
JA2 IPN 2PcP016 237 20081217 000439 20081217 010051	5230340	2009-05-28 14:30:06
JA2 IPN 2PcP026 003 20090317 003542 20090317 013155	4407820	2009-05-28 14:30:02
JA2 IPR 2PTP003 179 20080807 200321 20080807 205934	263672	2009-05-28 14:30:24
JA2 IPS 2PTP000 101 20080706 010301 20080706 015914	31909676	2009-05-28 14:30:34
JA2 IPS 2PTP000 104 20080706 035139 20080706 044752	31045968	2009-05-28 14:30:28
JA2 IPS 2PTP000 107 20080706 064018 20080706 073631	29734248	2009-05-28 14:30:33
JA2 IPS 2PTP000 109 20080706 083244 20080706 092857	33407064	2009-05-28 14:30:27
JA2 IPS 2PTP000 111 20080706 102510 20080706 112122	32390640	2009-05-28 14:30:32
JA2 IPS 2PTP000 112 20080706 112122 20080706 121735	32335732	2009-05-28 14:30:31
JA2 IPS 2PTP000 113 20080706 121735 20080706 131348	31155184	2009-05-28 14:30:25
JA2 IPS 2PTP000 117 20080706 160227 20080706 165840	33592956	2009-05-28 14:30:29
JA2 IPS 2PTP000 120 20080706 185106 20080706 194718	32237512	2009-05-28 14:30:26
JA2 IPS 2PTP000 122 20080706 204331 20080706 213944	34609380	2009-05-28 14:30:35
JA2 IPS 2PTP001 240 20080721 091523 20080721 101136	35451320	2009-05-28 14:30:36
JA2 IPS 2PTP001 243 20080721 120402 20080721 130015	35976008	2009-05-28 14:30:37
JA2 IPS 2PTP001 244 20080721 130015 20080721 135628	35647984	2009-05-28 14:30:38
JA2 IPS 2PTP001 245 20080721 135628 20080721 145241	35866604	2009-05-28 14:30:39
JA2 IPS 2PTP001 246 20080721 145241 20080721 154853	35374644	2009-05-28 14:30:40
JA2 IPS 2PTP001 247 20080721 154853 20080721 164506	35943056	2009-05-28 14:30:41
JA2 IPS 2PTP001 248 20080721 164506 20080721 174119	35680712	2009-05-28 14:30:42
JA2 IPS 2PTP001 249 20080721 174119 20080721 183732	35855608	2009-05-28 14:30:43
JA2 IPS 2PTP001 251 20080721 193345 20080721 202958	35516812	2009-05-28 14:30:44
JA2 IPS 2PTP001 252 20080721 202958 20080721 212611	36205400	2009-05-28 14:30:45
JA2 IPS 2PTP001 253 20080721 212611 20080721 222224	36052272	2009-05-28 14:30:46
JA2 IPS 2PTP001 254 20080721 222224 20080721 231836	35483860	2009-05-28 14:30:47
JA2 IPS 2PTP002 002 20080722 001449 20080722 011102	34248816	2009-05-28 14:30:49
JA2 IPS 2PTP002 003 20080722 011102 20080722 020715	32554352	2009-05-28 14:30:53
JA2 IPS 2PTP002 004 20080722 020715 20080722 030328	35495044	2009-05-28 14:30:54
JA2 IPS 2PTP002 005 20080722 030328 20080722 035941	35134068	2009-05-28 14:30:55
JA2 IPS 2PTP002 006 20080722 035941 20080722 045554	34981128	2009-05-28 14:30:56
JA2 IPS 2PTP002 008 20080722 055207 20080722 064819	36041500	2009-05-28 14:30:57
JA2 IPS 2PTP002 009 20080722 064819 20080722 074432	35888560	2009-05-28 14:31:04
JA2 IPS 2PTP002 011 20080722 084045 20080722 093658	35954052	2009-05-28 14:31:03
JA2 IPS 2PTP002 015 20080722 122537 20080722 132150	35888560	2009-05-28 14:31:02
JA2 IPS 2PTP002 017 20080722 141802 20080722 151416	35429364	2009-05-28 14:31:01
JA2 IPS 2PTP002 018 20080722 151416 20080722 161028	35177792	2009-05-28 14:31:00
JA2 IPS 2PTP002 021 20080722 180254 20080722 185907	35877376	2009-05-28 14:30:59
JA2 IPS 2PTP002 023 20080722 195520 20080722 205133	35440136	2009-05-28 14:30:52
JA2 IPS 2PTP002 024 20080722 205133 20080722 214745	36008548	2009-05-28 14:30:51
JA2 IPS 2PTP002 025 20080722 214745 20080722 224358	35680712	2009-05-28 14:30:50
JA2 IPS 2PTP002 026 20080722 224358 20080722 234011	33866072	2009-05-28 14:33:23
JA2 IPS 2PTP002 027 20080722 234011 20080723 003624	33002776	2009-05-28 14:33:22
JA2 IPS 2PTP002 029 20080723 013237 20080723 022850	33527464	2009-05-28 14:33:22
JA2 IPS 2PTP002 032 20080723 042116 20080723 051729	34981128	2009-05-28 14:33:21
JA2 IPS 2PTP002 034 20080723 061341 20080723 070954	35768116	2009-05-28 14:31:34
JA2 IPS 2PTP002 035 20080723 070954 20080723 080607	35844836	2009-05-28 14:33:19
JA2 IPS 2PTP002 036 20080723 080607 20080723 090220	35090344	2009-05-28 14:33:25
JA2 IPS 2PTP002 038 20080723 095833 20080723 105446	33374336	2009-05-28 14:33:24
JA2 IPS 2PTP002 039 20080723 105446 20080723 115059	35495044	2009-05-28 14:35:27
JA2 IPS 2PTP002 040 20080723 115059 20080723 124712	35582492	2009-05-28 14:35:28

JA2 IPS 2PTP002 042 20080723 134325 20080723 143937	35833652	2009-05-28 14:35:29
JA2 IPS 2PTP002 045 20080723 163203 20080723 172816	36041500	2009-05-28 14:35:30
JA2 IPS 2PTP002 048 20080723 192042 20080723 201655	36172672	2009-05-28 14:35:30
JA2 IPS 2PTP002 050 20080723 211308 20080723 220920	35986780	2009-05-28 14:35:31
JA2 IPS 2PTP002 052 20080723 230533 20080724 000146	34740552	2009-05-28 14:35:32
JA2 IPS 2PTP002 053 20080724 000146 20080724 005759	33144720	2009-05-28 14:35:33
JA2 IPS 2PTP002 056 20080724 025025 20080724 034638	35232700	2009-05-28 14:35:34
JA2 IPS 2PTP002 057 20080724 034638 20080724 044251	35538768	2009-05-28 14:35:35
JA2 IPS 2PTP015 089 20081201 072621 20081201 082234	35789928	2009-05-28 14:35:37
JA2 IPS 2PTP015 090 20081201 082234 20081201 091847	35549540	2009-05-28 14:35:38
JA2 IPS 2PTP015 095 20081201 130339 20081201 135952	36128948	2009-05-28 14:35:39
JA2 IPS 2PTP015 098 20081201 155217 20081201 164830	36150904	2009-05-28 14:35:40
JA2 IPS 2PTP015 128 20081202 195844 20081202 205457	34084692	2009-05-28 14:35:47
JA2 IPS 2PTP015 137 20081203 042440 20081203 052052	36159887	2009-04-15 15:28:25
JA2 IPS 2PcP015 147 20081203 134648 20081203 144301	36039708	2009-05-28 14:35:41
JA2 IPS 2PcP015 150 20081203 163527 20081203 173140	36336140	2009-05-28 14:35:42
JA2 IPS 2PcP015 152 20081203 182753 20081203 192406	34832584	2009-05-28 14:35:42
JA2 IPS 2PcP015 155 20081203 211631 20081203 221244	34569232	2009-05-28 14:35:43
JA2 IPS 2PcP015 159 20081204 010123 20081204 015736	34042528	2009-05-28 14:35:44
JA2 IPS 2PcP016 170 20081214 091816 20081214 101429	34371812	2009-05-28 14:35:45
JA2 IPS 2PcP029 181 20090422 171728 20090422 181341	32396992	2009-05-28 14:35:46

3. GDR Data Corruption

The GDR datasets were generated and transmitted to CLASS in bulk, much later than the IGDRs. When a large volume of GDR data were being flowed through the ESPC FileManager, it was more likely that the corruption would occur. This forced us to modify the ‘GDR Blessing’ behavior in order to throttle the GDR data flows. Although this reduced the instances of data corruption, the problem was not eliminated until the software fix described in the introduction was finally implemented. We have analyzed all of the GDR data sent through the FileManager machine up until the software fix, and have identified the following files as having been affected by data corruption. All of these files have been replaced with the uncorrupted version:

Filename	Files Size	CLASS Availability
JA2 GPN 2PTP013 083 20081111 055202 20081111 064815	4998588	2009-05-15 17:12:35
JA2 GPN 2PTP016 067 20081210 084809 20081210 094422	5257228	2009-04-01 20:33:57
JA2 GPN 2PTP016 187 20081215 011355 20081215 021008	5200384	2009-04-01 20:31:01
JA2 GPS 2PTP005 054 20080822 185334 20080822 194946	36357992	2009-05-15 17:12:36
JA2 GPS 2PTP013 109 20081112 061337 20081112 070950	35820248	2009-05-15 17:12:37
JA2 GPS 2PTP013 132 20081113 034633 20081113 044246	35985004	2009-05-15 17:12:37
JA2 GPS 2PTP013 133 20081113 044246 20081113 053859	34821772	2009-05-15 17:12:38
JA2 GPS 2PTP013 153 20081113 232703 20081114 002316	35392684	2009-05-15 17:12:39
JA2 GPS 2PTP015 197 20081205 123732 20081205 133345	36303060	2009-04-03 14:47:33
JA2 GPS 2PTP016 092 20081211 081331 20081211 090944	35622640	2009-04-01 20:31:04
JA2 GPS 2PTP016 106 20081211 212032 20081211 221644	35414408	2009-04-01 20:31:52
JA2 GPS 2PTP016 121 20081212 112345 20081212 121958	36007172	2009-04-01 20:31:05
JA2 GPS 2PTP016 150 20081213 143358 20081213 153011	36346952	2009-04-01 20:31:49
JA2 GPS 2PTP016 205 20081215 180546 20081215 190159	31639696	2009-04-01 20:31:47
JA2 GPS 2PTP022 030 20090206 095924 20090206 105537	35194948	2009-05-15 17:12:40

4. OGDR Data Corruption

The OGDR datasets produced by NOAA and EUMETSAT do not include a large waveform product like the S-GDR/S-IGDR; in addition to the OGDR and OGDR-SSHA files (both in NetCDF format) there is a third near real-time data set, the OGDR-BUFR, which is in BUFR format. The OGDR-BUFR files are distributed via the WMO/GTS network and EUMETSAT's satellite-broadcast system EUMETCAST. The analysis of the OGDR products for cycles 0-31 is complete, and NO files currently at CLASS have been found to be corrupt. One known instance of OGDR file corruption was immediately remediated. This affected the OGDR:

Filename	Files Size	CLASS Availability
JA2_OPN_2PcS017_013_20081218_050107_20081218_065921	10611784	2008-12-22 22 17:18:15

5. Summary

A file corruption analysis of all the level-2 Jason-2 data sets at CLASS has been completed. For the near real-time OGDR family no cases of corruption were found in the CLASS archive; one OGDR that suffered corruption was replaced immediately upon detection in Dec, 2008. For the intermediate IGDR family the analysis identified 30 files which had been reprocessed at CNES but not properly archived at CLASS. The version at CLASS was older and in general significantly smaller. The IGDR analysis also identified an additional 72 files that had been corrupted by the 'chown' bug. All of the corrupt/small files were replaced by 14:35 UTC on 2009-05-28. For the highest-quality GDR family, 15 files were identified that had been corrupted by the 'chown' bug and all had been replaced by 17:12 UTC on 2009-05-15. The tables above list the filesizes and dates of availability at CLASS for all of the uncorrupted replacement files. Finally, the bug that led to the file corruption has been identified and repaired in the ESPC FileManager code. No further instances of this corruption are anticipated.